Privacy Evaluation of Biohashing Methods

Dr. Cagatay KARABAT
TUBITAK BILGEM UEKAE
The Scientific and Technological Research Council of Turkey

International Biometric Performance Testing Conference 2014
Gaithersburg, April 1-3 2014

Agenda

- Privacy for Biometrics
- What is Biohashing?
- Privacy Evaluation Framework
- Limitations of the Metrics
- Simulation Results
- Conclusion

Privacy for Biometrics

Personal Data:

Any information relating to **an identified or identifiable natural person**; an identifiable person is one who can be identified, directly or indirectly, ... by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. [Directive 95/46/EC]

Problem:

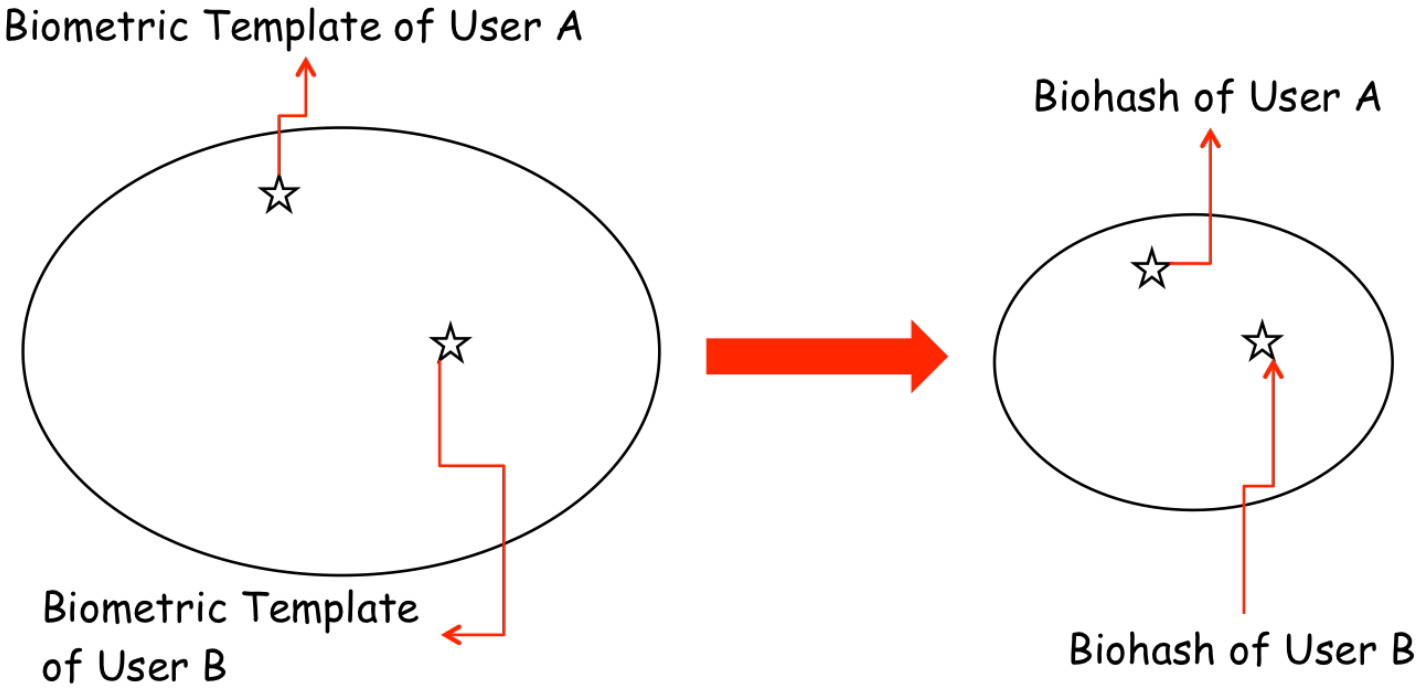
- Increasing usage of **biometric** data & biometric systems
- **Cannot** be **revoked** & reissued.

Threats:

- Biometric → reveal **sensitive & private** information
  skin color, age, sex, ethnic origin etc.
- Cross-matching of biometric traits → linkability, tracking, profiling
- In case of compromise → Identity theft.

What is biohasing?

- Biohash is a short and pseudo-random representation of biometric itself.
- It is an irreversible compressed representation of biometric data generated by using a secret key.



Verification via biohashes…

Biohash Vector

Enrollment

Stage

Authentication

Verify ?

Authentication

Stage

Biohash Vector

Privacy Evaluation Framework

Threat Models
- Naïve Model
- Advance Model

Protection Goals
- Diversification
- Unlinkability
- Privacy Leakage

Evaluation Metrics
- Entropy
- Conditional Entropy
- Distance Measure

Threat Models

- Threat Models
  - Naïve Model
  - Advanced Model

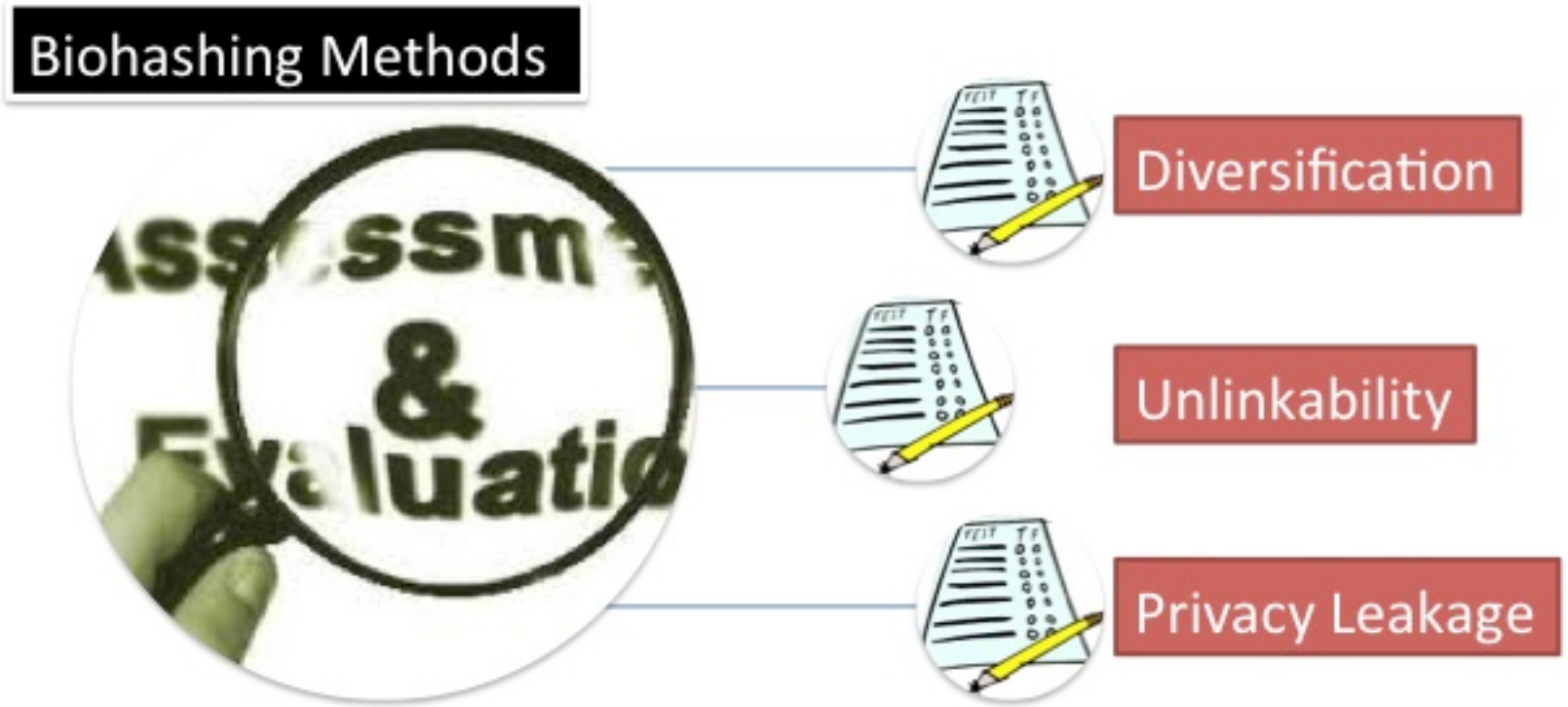
Length of Biohash Vector	Equal Error Rate
512	0,158%
256	0,663%
128	1,215%

Equal Error Rates in Naive Threat Model for 128, 256, and 512 bit Biohash Vectors in BioSecure Database

Length of Biohash Vector	Equal Error Rate
512	13,124%
256	13,836%
128	15,025%

Equal Error Rates in Advance Threat Model for 128, 256, and 512 bit Biohash Vectors in BioSecure Database

Protection Goals



Protection Goals: Diversification

- Diversification is the **maximum** number of **independent** protected biometric template that can be generated from the **same biometric feature** of the user by a biometric template protection method.

- It can be measured by using entropy H(***B***).

Length of Biohash Vector	Entropy (bit)
512	510,8323
256	255,2518
128	127,6303

Protection Goals: Unlinkability

- Unlinkability between the biohash vector and the user means that these items of interest cannot be related with each other after adversary's observation.

$$d(\mathbf{B1};\mathbf{B2}) = \frac{1}{2}(H(\mathbf{B1}|\mathbf{B2}) + H(\mathbf{B2}|\mathbf{B1}))$$

- Distance measure satisfies the properties of a metric (triangle in-equality, non-negativity, indiscernibility and symmetry). This distance metric is also known as the Variation of information.

Protection Goals: Unlinkability

- ***Case 1:*** Attacker gets two biohashes, which are generated from the same key in different authentication sessions, of the same user. However, the attacker does not know the owner of the biohashes.

- ***Case 2:*** Attacker gets two biohashes, which are generated from the different keys in different authentication sessions, of the same user. However, the attacker does not know the owner of the biohashes.

Protection Goals: Unlinkability (Case 1)

Length of Biohash Vector	Distance Measure (bit)
512	350,1094
256	179,0823
128	94,6018

Simulation results in terms of bits for unlinkability in case user has single secret key in BioSecure face database - case 1

Length of Biohash Vector	Distance Measure (bit)
512	171,8002
256	85,9001
128	42,95

Simulation results in terms of bits for unlinkability in case user has single secret key in in FVC2002-DB1 database - Case 1

Length of Biohash Vector	Distance Measure (bit)
512	174,6066
256	87,3033
128	43,6517

Simulation results in terms of bits for unlinkability in case user has single secret key in in FVC2002-DB2 database - Case 1

Length of Biohash Vector	Distance Measure (bit)
512	176,3561
256	88,1780
128	44,0890

Simulation results in terms of bits for unlinkability in case user has single secret key in in FVC2002-DB3 database - Case 1

Protection Goals: Unlinkability (Case 2)

Length of Biohash Vector	Distance Measure (bit)
512	510,7575
256	254,3375
128	127,0781

Simulation results in terms of bits for unlinkability in case of user has multiple secret keys in BioSecure face database - case 2

Length of Biohash Vector	Distance Measure (bit)
512	243,3346
256	121,2642
128	60,4813

Simulation results in terms of bits for unlinkability in case user has multiple secret keys in in FVC2002-DB1 database – Case 2

Length of Biohash Vector	Distance Measure (bit)
512	243,2703
256	121,3589
128	60,4009

Simulation results in terms of bits for unlinkability in case user has multiple secret keys in in FVC2002-DB2 database – Case 2

Length of Biohash Vector	Distance Measure (bit)
512	243,2376
256	121,3628
128	59,9730

Simulation results in terms of bits for unlinkability in case user has multiple secret keys in in FVC2002-DB3 database – Case 2

Protection Goals: Privacy Leakage

- Privacy leakage quantifies **how much information** about biometric data contained in a binary biohash vector.

- The **probability distribution of biohash** plays a very important role in this privacy assessment. It is expected that a biohash, ***B***, has uniform distribution where a bit's probability being 1 or 0 is equal.

- On the other hand, the dependency of binary features is ignored in many biometric template protection methods. Thus, privacy preservation capability of these methods are highly overestimated.

Protection Goals: Privacy Leakage

Length of Biohash Vector	Privacy Leakage (bit)
512	105,7862
256	59,8323
128	30,6894

Simulation results in terms of bits for privacy leakage in BioSecure face database

Length of Biohash Vector	Privacy Leakage (bit)
512	132,9231
256	58,9703
128	31,9961

Simulation results in terms of bits for privacy leakage in FVC2002-DB1 database

Length of Biohash Vector	Privacy Leakage (bit)
512	131,7263
256	58,8425
128	30,3845

Simulation results in terms of bits for privacy leakage in FVC2002-DB2 database

Length of Biohash Vector	Privacy Leakage (bit)
512	97,0203
256	43,1812
128	22,8893

Simulation results in terms of bits for privacy leakage in FVC2002-DB3 database

Limitations of the Metrics

- One size does not fit all !

Limitations of the Metrics

- The probability estimation of biometric data, distribution or conditional distribution of biometric data or secrets might not be always possible due to high dimension of features, limited number of available biometric data.
- For template protection algorithms that are not based on information-theoretical security, mentioned metrics may not be suitable.
- For some of the template protection methods proposed in the literature, practical evaluations that depend on individual attacks can be used. With practical evaluations, a direct way to evaluate an algorithm by assessing the efficiency of a defined attack is obtained and what an adversary can achieve in practice can be simulated.
- Theoretical and practical evaluations are expected to complement each other

Conclusion

- **Face:** Biosecure ds2
- **Fingerprint:** FVC2002 DB1-DB2-DB3
- Evaluations are **independent from biometric data type**
- **Similar and comparable** results for both face and fingerprint

Protection Goal	Fullfillment Level
Diversification	Strong ☺
Unlinkability (case 1)	Weak ☹
Unlinkability (case 2)	Strong ☺
Privacy Leakage	Fair 😐

Acknowledgments

This work has been performed by the BEAT project 7th Framework Research Programme of the European Union (EU), grant agreement number: 284989. The authors would like to thank the EU for the financial support and the partners within the consortium for a fruitful collaboration. For more information about the BEAT consortium please visit http://www.beat-eu.org.

Thank You…